DPS HW Design Review

Randy Miller

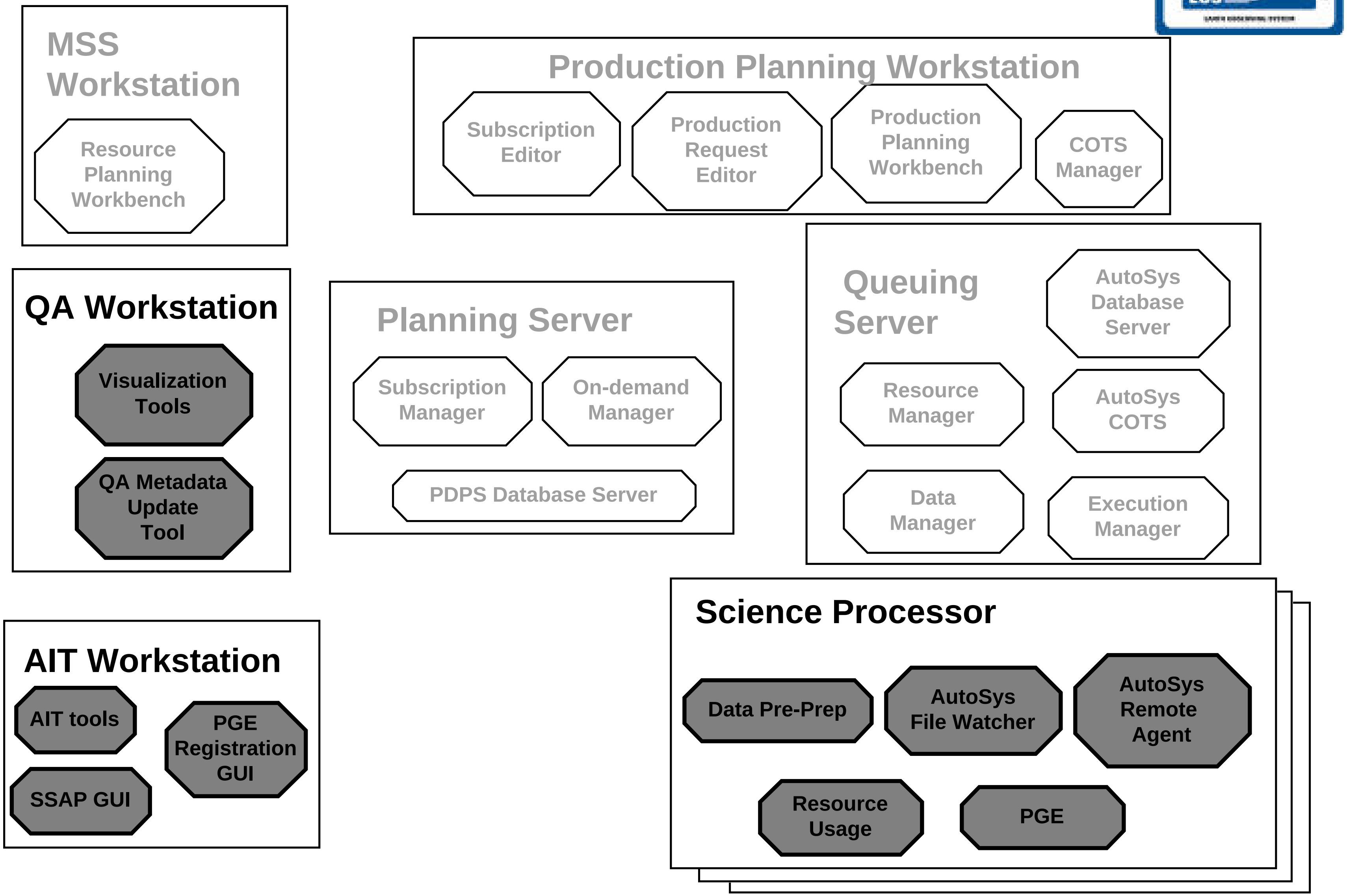
randallm@eos.hitc.com

22 April 1996

Overview

- Data Processing Subsystem (DPS) Hardware Configuration Items
- Requirements
- Sizing Analysis
- Specification
- Design Analysis
- Design Validation

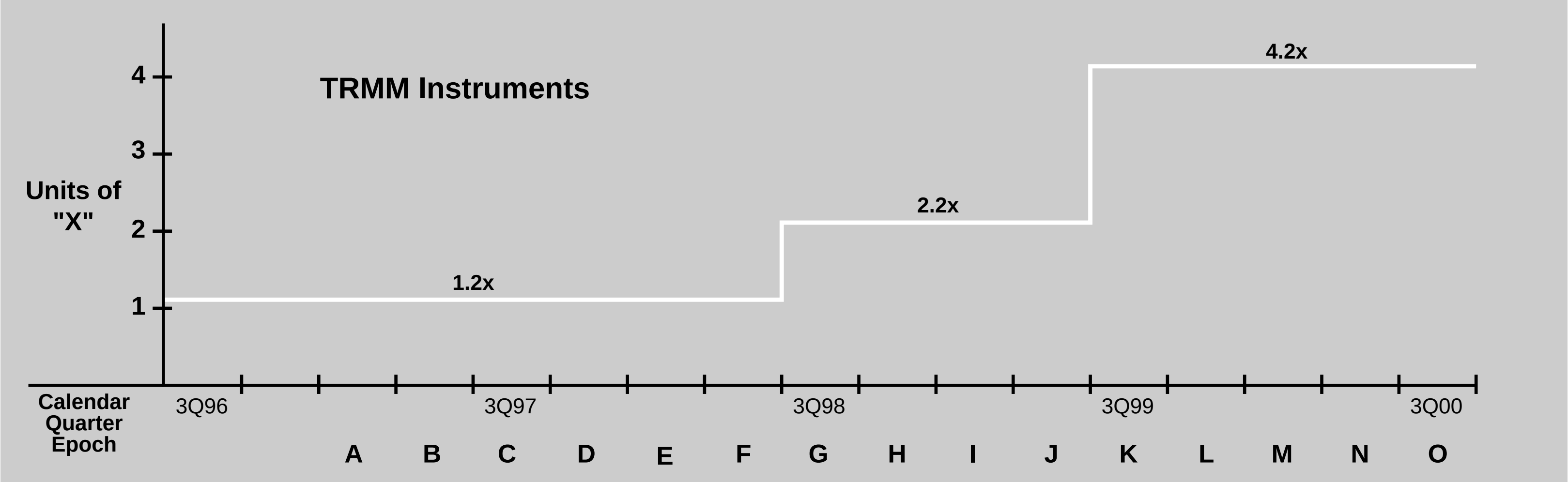
DPS Hardware Diagram

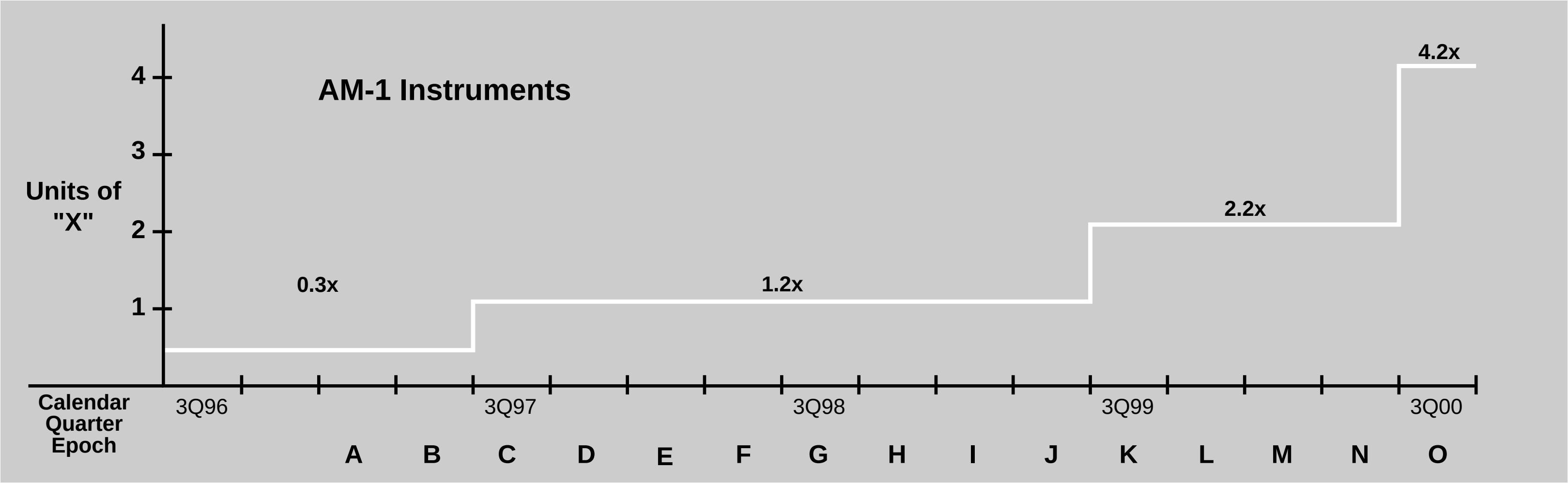


SPRHW F&PRS Requirements

- Timeliness
- Allocation of Resources
- Expandability
- Derating Of Processor Performance

SPRHW Phasing Requirements





Other SPRHW Requirements

- Selected Hardware Must Support Standard ECS Software
- Selected Hardware Must Support Highspeed Interconnect Protocol
  - HiPPI, OC-12 ATM, or FC-AL
- RMA Requirements
  - Availability of 96% or better
  - Mean Down Time not greater than 4 hours

Sizing Approach

- Static Modeling of AHWGP Inputs
  - Provides average loads for CPU, networks, disk I/O
- Dynamic Modeling
  - Provides more accurate loads for CPU, networks, disk I/O, and disk size
- Memory and I/O Survey
  - To provide memory requirements and refinements in I/O modeling
- Other Considerations And Adjustments

Static Modeling Analysis

- Process descriptions and volume timelines from the Ad Hoc Working Group for Production (AHWGP) are entered into a spreadsheet

- Processing requirements (CPU, network I/O, disk I/O, archive I/O) are summed by instrument and DAAC

- This approach supports several analyses:
  - Average load over long periods of time
  - Peak load on the worst case day
  - Loads under other assumptions (for example, spreading Level 3 production over multiple days)

Static Modeling Results (Summary)

		No. of Exec. /day	Processing (MFLOPS)	I/O Local to Processing (MB/sec)	Worst Case Processing I/O (MB/sec)	Worst Case Network I/O (MB/sec)	Best Case Deep Arch I/O (MB/sec)
ASTER	EDC	1,055	583.8	3.4	5.7	2.4	0.8
MODIS	EDC	4,920	1,051.0	28.8	63.7	35.0	12.6
DAO	GSFC	2	13,680.0	0.9	1.8	0.9	0.3
LIS	GSFC	2	2.0	0.1	0.2	0.1	0.0
MODIS	GSFC	19,328	4,712.9	125.5	243.8	118.3	13.8
DFA/MR	JPL	114	42.2	0.1	0.1	0.1	0.0
SWS	JPL	61	45.7	0.1	0.3	0.1	0.0
CERES(AM)	LaRC	103	1,826.6	3.8	7.6	3.8	3.3
CERES(TRMM)	LaRC	103	894.7	1.5	3.0	1.5	1.0
MISR	LaRC	566	3,299.0	18.6	26.5	8.0	2.7
MOPITT	LaRC	4	9.4	0.1	0.2	0.1	0.0
SAGE	LaRC	1	3.4	0.0	0.0	0.0	0.0
MODIS	NSIDC	1,705	14.8	0.4	0.9	0.4	0.2

Dynamic Modeling Analysis

- Event Driven Simulation Implemented Using BONeS
  - Models execution of each PGE
  - Models archiving of each granule
  - Models each user pull
- Driven By Technical Baseline And System Design
- Outputs Include
  - Resource utilization versus time
  - Queue depth over time for each resource
  - Time-averaged resource utilizations
  - Elapsed time for events (e.g., PGE turn-around times)

Dynamic Modeling Status

- Now Using February 1996 Technical Baseline To Establish Push Load
- Initial Runs Were Done Without MODIS Level 3s
  - Baseline simulations for EDC(ASTER), LaRC, NSIDC, JPL
  - Failover simulations for EDC(ASTER), LaRC
  - 2X, 4X, and 10X User Pull simulations
- Currently Modeling Execution Of Tile-Oriented MODIS Level 3s
  - Tile-oriented PGEs execute in batches
  - Results are being reviewed from full system baseline simulation including MODIS Level 3s

Memory And I/O Survey

- The objective of the survey was to gather the best possible data stating memory requirements and disk I/O characteristics for each PGE, to support system sizing

- The responses received to date have been minimal:
  - ASTER provided data for essentially all of their PGEs
  - CERES, MODIS, MISR, and DAO data unavailable at this time

- For the CDR design, assumptions were made
  - 128 MB of RAM per processor
  - 8 MB/sec per SCSI-2 channel/controller

- Assumptions will be validated via SSI&T and benchmarking

Other Considerations and Adjustments

- Additions Of CPU For Overheads (e.g., Network I/O)
- Addition Of Resources To Meet RMA Requirements
- Re-Use And/Or Upgrade Of Existing Equipment
- Configuration Of Disk Groupings For Striping
- Extrapolation From The Old Baseline To The New Baseline Where Modeling Results Are Not Yet Available
- Rounding Up To The Next Configurable Increment

SPRHW Specification

- Top-Level Summary

- Detailed System Specifications

- Specifications By Component Type
  - CPU
  - Memory
  - Disk
  - Network
  - Enclosure

SPRHW Top-Level Summary

						Epoch K (3Q99)			
Site	String	System	Derated Processing [MF]	RAM [MB]	RAM [interleave]	Disk Channels [N]	Staging I/O [MB/s]	Processing I/O [MB/s]	Net Disk Space [GB]
EDC	AI&T	-1	1,375	1,024	8	10	0.0	0.0	720
	ASTER	-4	1,375	1,024	4	2	3.9	5.3	110
	MODIS	-5	1,375	1,024	8	10	32.4	24.1	720
		-6	1,375	1,024	8	10	32.4	24.1	720
		Total	2,750	2,048	16	20	64.8	48.2	1,440
EDC	All	All	5,500	4,096	28	32	68.7	53.5	2,270
GSFC	AI&T	-1							
		-9	1,100	1,024	8	8	0.0	0.0	247
		-10							
		Total	1,100	1,024	8	8	0.0	0.0	247
	LIS & COLOR	-4	75	512	2	2	0.3	0.3	68
	MODIS	-1	2,475	2,048	8	10	8.0	60.4	247
		-5	2,475	2,048	8	10	8.0	60.4	247
		-6	2,475	2,048	8	10	8.0	60.4	247
		-8	2,475	2,048	8	10	8.0	60.4	247
		-11							
		-12							
		Total	9,900	8,192	32	40	32	242	989
GSFC	All	All	11,075	9,728	42	50	32.3	241.9	1,304
JPL	AI&T	-1	137	128	1	1	0.0	0.0	17
	DFA/MR & SWS	-2	825	512	2	1	1.8	1.8	17
JPL	All	All	962	640	3	2	1.8	1.8	34
LaRC	AI&T	-1	1,925	2,048	8	4	0.0	0.0	288
		-13							
		Total	1,925	2,048	8	4	0.0	0.0	288
	CERES TRMM	-5	1,080	2,048	8	2	1.8	2.3	69
		-6	1,620	2,048	8	2	1.8	2.3	103
		Total	2,700	4,096	16	4	3.6	4.6	172
	CERES AM-1	-8	2,200	2,048	8	3	3.5	3.8	432
		-11	2,200	2,048	8	3	3.5	3.8	432
		Total	4,400	4,096	16	6	7.0	7.6	864
	MISR	-9	3,300	2,048	8	4	5.4	18.6	288
		-10	3,300	2,048	8	4	5.4	18.6	288
		-12							
		Total	6,600	4,096	16	8	10.8	37.2	576
LARC	All	All	15,625	14,336	56	22	21.4	49.4	1,900
NSIDC	AI&T	-1	137	128	1	1	0.0	0.0	17
	MODIS	-2	550	512	2	1	4.0	4.0	17
NSIDC	All	All	687	640	3	2	4.0	4.0	34
All	All	All	33,849	29,440	132	108	128.2	350.6	5,542

SPRHW Detail (Example)

C	G	K
3Q97	3Q98	3Q99
SPRHW-EDC-4	**SPRHW-EDC-4**	**SPRHW-EDC-4**
Function: EDC AI&T	Function: ASTER	Function: ASTER
Cabinet: Power Challenge XL	Cabinet: Power Challenge XL	Cabinet: Power Challenge XL
Console: Character	Console: Character	Console: Character
CPU: 6 x 275 MHz R10000	CPU: 6 x 275 MHz R10000	CPU: 10 x 275 MHz R10000
RAM: 1 GB/4-way interleaved	RAM: 1 GB/4-way interleaved	RAM: 1 GB/4-way interleaved
IO4: Two	IO4: Two	IO4: Two
HIO-1 (1,1): FDDI	HIO-1 (1,1): FDDI	HIO-1 (1,1): FDDI
HIO-2 (1,2): SCSI	HIO-2 (1,2): SCSI	HIO-2 (1,2): SCSI
HIO-3 (2,1): HiPPI	HIO-3 (2,1): HiPPI	HIO-3 (2,1): HiPPI
HIO-4 (2,2): Unused	HIO-4 (2,2): Unused	HIO-4 (2,2): Unused
SCSI-0 (1,0,1): CD-ROM	SCSI-0 (1,0,1): CD-ROM	SCSI-0 (1,0,1): CD-ROM
SCSI-1 (1,0,2): Two 4.3 GB Internal Disks	SCSI-1 (1,0,2): Two 4.3 GB Internal Disks	SCSI-1 (1,0,2): Two 4.3 GB Internal Disks
SCSI-2 (1,2,1): RAID-1 SP1	SCSI-2 (1,2,1): RAID-1 SP1	SCSI-2 (1,2,1): RAID-1 SP1
SCSI-3 (1,2,2): RAID-1 SP2	SCSI-3 (1,2,2): RAID-1 SP2	SCSI-3 (1,2,2): RAID-1 SP2
SCSI-4 (1,2,3): Unused	SCSI-4 (1,2,3): Unused	SCSI-4 (1,2,3): Unused
RAID-1: 10 x 9 GB RAID 5 (Cabinet 1)	RAID-1: 10 x 9 GB RAID 5 (Cabinet 1)	RAID-1: 15 x 9 GB RAID 5 (Cabinet 1)

SPRHW Processors

- SGI Power Challenge Processors

  - R10000
    - For new machines purchased for Release B
    - Currently shipping 200 MHz chips; 275 MHz chips announced
    - Two floating point operations per clock cycle
    - Two or four processors per board; up to 36 processors per system

  - R8000
    - Retained from Release A for CERES TRMM processing; other Release A R8000s traded in for R10000 processors
    - Retaining only 90 MHz processors
    - Four floating point operations per clock cycle
    - Two processors per board; up to 18 processors per system

SPRHW Processors (cont.)

- SGI Power Challenge Processors (Continued)
  - R4600
    - Retained from Release A in one system to support LIS and COLOR
    - Retaining only 150 MHz processors
    - One floating point operation per two clock cycles
    - One, two, or four processors per board; up to 36 processors per system
- Number of CPUs per system configured to satisfy DAAC and instrument requirements
  - Up to 20 CPUs per system at Epochs C and G (Initial purchase)
  - Up to 24 CPUs per system at Epoch K (Second purchase)
  - Some small systems configured at NSIDC, JPL, GSFC (LIS & COLOR)

Random Access Memory

- Approximately 128 MB per processor:
  - 128 MB for uniprocessors (NSIDC and JPL AI&T/AQA systems)
  - 512 MB for 4 processors (NSIDC, JPL, and LIS/COLOR systems)
  - One GB for 5 to 12 processors (EDC, GSFC AI&T systems)
  - Two GB for 13 or more processors (LaRC, GSFC systems)
- Memory Interleaving
  - Smaller systems 1 or 2 way interleaved
  - 1 GB x 4 or 8 way (Depending on I/O expected on system)
  - 2 GB x 8 way

I/O Subsystems

- The Challenge architecture supports up to 6 I/O subsystems (IO4 cards) per system
- Each IO4 supports up to 320 MB/s
- Each IO4 provides two FWD SCSI-2 channels and two HIO ports
- Each HIO can support
  - A card with three FWD SCSI-2 channels, or
  - A HiPPI connection, or
  - A FDDI connection
- The number of IO4 cards per system is driven by the number of connections (HiPPI + FDDI + SCSI-2) required

Internal Disk Storage

- Used for
  - Swap Space
  - Operating System
  - COTS and ECS Software
- Sized as four times the RAM size, plus two GB
- Configured as one, two, or three 4.3 or 9 GB disks, on a single SCSI-2 channel

External Disk Arrays

- **Each system’s external array is sized according to the I/O rates and storage requirements for its intended instrument:**
  - **Some instruments have a high “size to rate” ratio (CERES AM-1); they need many disks but few controllers/channels**
  - **Some instruments have a low “size to rate” ratio (MODIS at GSFC); they require more controllers and fewer disks/channels**

- **Arrays are specified as SGI RAID 5 (SCSI-2 based)**
  - **Waiting to see new Fibre Channel offerings under SGI “Gold Seal” program**

Network Interfaces

- HiPPI
  - At LaRC, GSFC, and EDC, each SPRHW system (except LIS/COLOR) will have one HiPPI interface
- FDDI
  - Each SPRHW system will have one FDDI interface

Failure Recovery

What Fails, And For How Long?

- Science Software - Failure could be extended, but is likely to have limited impact
- SPRHW Hardware
  - Redundant Components (CPU, Memory) - On call vendor maintenance should limit outage to a single shift; system may function in degraded mode
  - Single Points Of Failure - On call vendor maintenance should limit outage to a single shift, but system will be unavailable during duration of failure
- Other Subsystems/External Systems - Failure could be extended, with broad impact on first-time processing, but possibly less impact on reprocessing

Expandability

- Is There Margin Above Requirements?

- How Far Can The Planned Boxes Be Expanded?
  - Up to 36 CPUs per system
  - Up to 16 GB of RAM per system
  - Up to 40 SCSI-2 channels per system

- Where Will Upgrades Take Us?
  - Faster CPUs
  - New disk technologies (Fibre Channel)

- Expansion By Adding Boxes

AITHW

Function: The function of AITHW is to support the integration and test of science software at the DAAC. AITHW provides tools (code management, debugging, performance) for software integration and test, and seats (development stations) for the I&T team. Remote access to the AI&T tools is also provided to the instrument teams.

Specification: At each processing DAAC, AITHW provides a tools server (a Sun 20/50 with 128 MB of RAM and 4 GB of disk) and a number of developer's stations (Sun 20/50 workstations and/or NCD X-terminals). A target environment (SGI compute platform) for AI&T is provided in the sizing of SPRHW.

AQAHW

Function: The function of AQAHW is to provide the DAAC with resources to perform non-science Quality Assurance testing.

Specification: At each processing DAAC, AQAHW is provided as an SGI visualization workstation (SGI Indigo 2 IMPACT 10000 workstation, with 128 MB RAM and approximately 17 GB of disk space).